

The pre-assessment learning effects of consequential assessment: modelling how the examination game is played

Citation for published version (APA):

Cilliers, F. J. (2012). *The pre-assessment learning effects of consequential assessment: modelling how the examination game is played*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20120905fc>

Document status and date:

Published: 01/01/2012

DOI:

[10.26481/dis.20120905fc](https://doi.org/10.26481/dis.20120905fc)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

Googling the exact phrase “assessment drives learning” in early 2012 yields 8 800 results, of which 610 are scholarly articles. This may be a rather unscientific way to start the summary of a dissertation, but it does give some indication of just how widely held this belief is. Yet although it is widely professed and a search of the literature on the topic reveals that there are vast numbers of research papers about various aspects of the topic, there is little uniformity about exactly what is meant by “assessment drives learning”. There are also no validated models explaining the relationship between assessment and learning. It is therefore perhaps not surprising that there is a dearth of success stories in the literature about purposively using assessment to drive learning in desired ways. The ultimate goal that inspired this research, then, is the hope that has since been expressed by Krupat and Dienstag (2009) thus: that we should be able to utilise “approaches and methods of assessment that fulfil their full potential to reinforce our educational goals” (p548). The specific issue that this dissertation is concerned with is deriving and starting to validate a model of the pre-assessment learning effects of summative assessment.

In **Chapter One**, the nature of the literature about the relationship between assessment and learning is reviewed and problematized. The lack of a mechanism and model explaining the relationship between assessment and pre-assessment learning effects is highlighted. The delimitation of the dissertation to the topic of the pre-assessment learning effects of assessment is explained and the fourfold goals of the dissertation spelled out i.e., 1) to discern what components of assessment influenced what dimensions of learning in an authentic, complex system of summative assessment; 2) to explore what mechanism was at play at the interface between assessment and learning in this assessment system; 3) to construct a theoretically grounded model and finally 4) to take steps towards validating this model.

Chapter Two makes the argument that it has become axiomatic that assessment impacts powerfully on student learning, but that there is a surprising dearth of research on how this happens. A study is described that explored the mechanism of impact of summative assessment on the process of learning of theory in higher education. Individual, in-depth interviews were conducted with medical students and analysed qualitatively. The impact of assessment on learning was found to be mediated through various determinants of action. Respondents’ learning behaviour was influenced by: appraising the impact of assessment; appraising their learning response; their perceptions of agency; and interpersonal factors such as normative beliefs and motivation to comply with those beliefs. The results of the study add to scant extant evidence about the impact of assessment on learning and lead to the proposal of a mechanism to explain this impact.

Chapter Three highlights and reviews the surprisingly small amount of research that has been published emanating from authentic higher education and health sciences

education settings about the nature and mechanism of the pre-assessment learning effects of summative assessment. This study explored the pre-assessment learning effects of summative assessment in theoretical modules by exploring the variables at play in a multifaceted assessment system and the relationships between them. Using a grounded theory strategy, in-depth interviews were conducted with individual medical students and analysed qualitatively. Respondents' learning was influenced by task demands and system design. Assessment impacted on respondents' cognitive processing activities and metacognitive regulation activities. Individually, our findings confirm findings from other studies in disparate non-medical settings and identify some new factors at play in this setting. Taken together, findings from this study provide, for the first time, some insight into how a whole assessment system influences student learning over time in a medical education setting. The findings from this authentic and complex setting paint a nuanced picture of how intricate and multifaceted interactions between various factors in an assessment system interact to influence student learning. A model linking the sources, mechanism and consequences of the pre-assessment learning effects of summative assessment is proposed that could help enhance the use of summative assessment as a tool to augment learning. The model is depicted on page 120 of this dissertation.

Chapter Four addresses the issue that no validated model exists to explain the learning effects of assessment and how this is a problem when designing and re-searching assessment for learning. The challenge addressed in this study is how to validate the model described in Chapter Three. The purpose of this study was to explore whether the model was operational in a clinical context as a first step in this process. Given the complexity of the model, a qualitative approach was adopted. Data from in-depth interviews with eighteen medical students were subject to content analysis. A code book developed previously using grounded theory was utilised. During analysis, we remained alert to data that might not conform to the coding framework and open to the possibility of deploying inductive coding. The three components of the model i.e., assessment factors, mechanism factors and learning effects were all evident in the clinical context. Associations between these components could all be explained by the model. "Interaction with preceptors" was identified as a new subcomponent of assessment factors. The model could explain the interrelationships of the three facets of this subcomponent i.e., regular accountability, personal consequences and emotional valence of the learning environment, with previously described components of the model. The model could thus be utilized to analyse and explain observations in an assessment context different to that from which it was derived. In the clinical setting, the (negative) influence of preceptors on student learning was particularly prominent. In this setting, learning effects resulted not only from the high-stakes nature of summative assessment but also from personal stakes, e.g. for esteem and agency. The results suggest that to influ-

ence student learning, consequences should accrue from assessment that are immediate, concrete and substantial. The model could have utility as a planning or diagnostic tool in practice and research settings. The proposal that the model is one of consequential assessment rather than summative assessment is discussed.

In **Chapter Five**, we review how we previously developed and started validating a model of the pre-assessment learning effects of consequential assessment. The purpose of this study was to continue the process of validation. As the most stringent test of the model, the aims were to determine whether the least common associations between assessment factors and learning effects in the model and the components involved were testable and whether support could be found for the role of this subset of individual components and the associations between them in a broader but similar population to that in which the model was initially derived. Twenty-one different associations were investigated in a cross-sectional survey using a purpose-made questionnaire. Each association was tested using a pair of situational test items that were developed based on a grounded theory and structured using the logic of the key-features approach. For each association, the chi-square statistic was used to determine whether there were significant associations between assessment factors and learning effects. An adjusted alpha was used to assign significance. The frequency of involvement of mechanism factors was calculated. Almost all associations and components were testable. There were significant associations between assessment factors and learning effects for 13/21 of this weakest set of associations being investigated, even taking $p < 0.00625$ as significant. The role of all eight learning effects, all ten mechanism factors and six out of eight assessment factors investigated were substantiated. For the least common subset of associations in the model, the role of most components involved and the associations between them are therefore supported. This bodes well for future studies of the commonest associations in the model. The model also holds in a broader but similar population to that from which it was derived. Results illustrate how a range of factors can confound the results of interventions aimed at using assessment to influence learning.

Chapter Six revisits the rationale for model validation i.e., that interventions based on a validated model should produce better outcomes than efforts based on weaker theory. Despite extensive literature on the topic, no validated model explaining the learning effects of assessment exists. Before using a recently developed model to inform interventions, it needs to be validated. The purpose of this study was to explore whether tentative generalizability findings related to the model could be replicated in novel assessment contexts. Cross-sectional surveys of 419 students at two other universities were undertaken, using a purpose-made questionnaire comprising written situational tests. For stringency, the model's 21 least common asso-

ciations between assessment factors and learning effects were investigated. Thirteen of these associations were significant ($p < 0.00625$). The role of seven of eight assessment factors, seven of eight learning effects and all ten mechanism factors investigated, were substantiated. Four mechanism factors (agency, response efficacy, response value and impact appraisal) together mediated most associations. This further support for many of the model's least common associations again bodes well for future studies of the commoner associations. Although model validation is an on-going process, these results move the model one step closer to the stage of usefully informing interventions.

Chapter Seven presents an overview of findings and conclusions from Chapters 2 – 6. Chapter 2 resulted in the proposal of a novel mechanism to explain why students respond to assessment in the ways that they do and not merely what their response is. Findings from Chapter 3 lead to the proposal of a model of the pre-assessment learning effects of summative assessment. Based on the findings reported in Chapters 4 – 6, the conclusion is drawn that evidence collected thus far supports the validity of the model in various settings and that the model can be considered one of consequential assessment rather than of summative assessment.

Four main findings from this work are then discussed: 1) associations between assessment factors and learning effects are complex and contextual; 2) a novel mechanism is in play in theory and clinical assessment contexts; 3) the model applies to consequential assessment rather than just summative assessment and 4) a model has been devised for which there is validity evidence. After expanding on each of these four findings, existing evidence for the validity of the model is reviewed in light of a hierarchical set of criteria proposed by Prochaska, Wright & Velicer (2008).

The limitations of this work are discussed. For the initial studies, these include limitations inherent in qualitative research. They also include the limitations inherent in utilising volunteers for both qualitative and quantitative research. For the quantitative work, the potential for generalizability of findings is as yet limited, given that only the least common associations in the model have been studied in a limited range of assessment contexts thus far.

The strengths of this dissertation are that the mechanism and model are derived from the lived experiences of students across the span of their degree studies, in a complex, authentic setting. The qualitative work also addressed many of the threats to causal inference identified by Maxwell (2004). Overall, the dissertation proposes a model explaining the pre-assessment learning effects of consequential assessment for which validation has been undertaken. This makes a contribution on two fronts. Firstly, the model offers a new perspective on how to relate assessment and pre-assessment learning by linking assessment factors and learning effects by way of a

novel mechanism. Secondly, no other model proposed in the past has – as yet, anyway – been validated to any extent.

An agenda for future research is discussed before, finally, the implications of the research for teaching and educational development practice are explored.

References

- Krupat, E., & Dienstag, J. L. (2009). Commentary: Assessment Is an Educational Tool. *Academic Medicine*, 84(5), 548-550.
- Maxwell, J. A. (2004). Using Qualitative Methods for Causal Explanation. *Field Methods*, 16(3), 243-264.
- Prochaska, J. O., Wright, J. A., & Velicer, W. F. (2008). Evaluating Theories of Health Behavior Change: A Hierarchy of Criteria Applied to the Transtheoretical Model. *Applied Psychology*, 57(4), 561-588.

Samenvatting
Summary in Dutch

Zoeken op het internet naar de uitdrukking “toetsing stuurt leren” leverde, begin 2012, 8800 resultaten op, waaronder 610 wetenschappelijke artikelen. Al lijkt dit een onwetenschappelijke opening van de samenvatting van een proefschrift, het illustreert hoe breed de opvatting “toetsing stuurt leren” gedragen wordt. Des te opmerkelijker is het ontbreken - ondanks het indrukwekkende aantal wetenschappelijk publicaties dat diverse aspecten van dit idee belicht - van gevalideerde modellen die een verklaring bieden voor de veronderstelde relatie tussen toetsing en leren. Het wekt dan ook geen verwondering dat er in de literatuur weinig melding wordt gemaakt van succesvolle toepassingen van toetsing als methode om gewenste vormen van leren te stimuleren. Het onderzoek in dit proefschrift is ingegeven door de wens uitgesproken door Krupat en Dienstag (2009) dat ‘toetsbenaderingen en -methoden de verwachting zullen waarmaken dat ze een bijdrage kunnen leveren aan de realisatie van onze onderwijsdoelen ’ (p. 548). Het onderzoek in dit proefschrift had tot doel een model te ontwikkelen van het effect van summatieve toetsing op het voorafgaande leren en daarnaast is een begin gemaakt met het valideren van dit model.

In **Hoofdstuk 1** wordt een overzicht gegeven van de literatuur over de relatie tussen toetsing en leren en worden problemen van deze relatie belicht. Hierbij ligt de nadruk op het ontbreken van een verklarend mechanisme of model. Toegelicht wordt waarom het onderzoek in dit proefschrift uitsluitend gericht is op de invloed van toetsing op voorafgaande leereffecten en daarnaast worden de vier onderzoeksdoelen beschreven: 1) bepalen welke onderdelen van toetsing van invloed zijn op welke dimensies van het leren in een authentiek, complex systeem van summatieve toetsing; 2) onderzoeken welk mechanisme in dit toetssysteem een rol speelt op het raakvlak tussen toetsen en leren; 3) ontwikkelen van een theoretisch goed onderbouwd model en ten slotte 4) een eerste validering van het ontwikkelde model.

In **Hoofdstuk 2** wordt beargumenteerd dat weliswaar algemeen als axioma wordt aanvaard dat toetsing het leren van studenten sterk beïnvloedt, maar dat er een opmerkelijk gebrek is aan onderzoek naar de wijze waarop dit effect tot stand komt. Er is gezocht naar een antwoord op de vraag welk mechanisme ten grondslag ligt aan het effect van summatieve toetsing op theoretisch leren in het hoger onderwijs. Hiertoe zijn diepgaande interviews met individuele geneeskundestudenten afgenomen en geanalyseerd. De resultaten lieten zien dat verschillende factoren het handelen van de deelnemers bepaalden en daarmee indirect een rol speelden bij de totstandkoming van het effect van toetsing op leren. Het leergedrag werd beïnvloed door: de inschatting van het belang van toetsing; de inschatting van de eigen leerrespons; het gepercipieerde vermogen tot zelfstandig handelen; en sociale factoren, zoals normatieve opvattingen en de motivatie om daarnaar te handelen.

De resultaten van dit onderzoek vormen een aanvulling op het weinige bewijsmateriaal dat beschikbaar is ter ondersteuning van het effect van toetsing op leren. Daarnaast onderbouwen ze het mechanisme dat voorgesteld wordt als verklaring van dit effect.

Hoofdstuk 3 geeft een overzicht en benadrukt het geringe aantal publicaties waarin onderzoek gerapporteerd wordt dat is uitgevoerd in authentieke leeromgevingen in het hoger en gezondheidszorgonderwijs naar de manier waarop summatieve toetsing het voorafgaande leren bepaalt en het onderliggende mechanisme. In een exploratief onderzoek is onderzocht welke variabelen, met hun onderlinge relaties, in een toetssysteem met veel verschillende facetten een rol spelen bij de totstandkoming van effecten van summatieve toetsing van theoretische modules op het voorafgaande leren. Het onderzoek werd uitgevoerd volgens de 'grounded theory'-methode en bestond uit afname en kwalitatieve analyse van interviews met individuele geneeskundestudenten. Het leren van de deelnemers werd beïnvloed door taakgerelateerde eisen en de aard van het toetssysteem. Toetsing beïnvloedde de activiteiten van de deelnemers met betrekking tot cognitieve verwerking en metacognitieve regulering. De resultaten bevestigen bevindingen uit ander onderzoek in verschillende niet-medische omgevingen en brengen nieuwe factoren aan het licht die daar een rol spelen. Dit is het eerste onderzoek dat inzicht biedt in de wijze waarop het toetssysteem als geheel het leren van studenten gedurende het hele curriculum beïnvloedt. De resultaten van dit onderzoek in een authentieke en complexe leeromgeving schetsen een genuanceerd beeld van de manier waarop ingewikkelde en veelvormige interacties tussen verschillende factoren in het toetssysteem het leren van studenten beïnvloeden. Een model wordt voorgesteld waarin bronnen, mechanisme en gevolgen van het effect van summatieve toetsing op leren met elkaar in verband worden gebracht met als doel summatieve toetsing optimaal te kunnen toepassen als instrument om het leren te bevorderen. Het model is te vinden op bladzijde 120 van dit proefschrift.

Hoofdstuk vier gaat in op het ontbreken van een gevalideerd model dat de leereffecten van toetsing kan verklaren en hoe dit een belemmering vormt voor het ontwerpen en onderzoeken van toetsing als stimulans voor leren. Gezien de complexiteit van het model is een kwalitatieve onderzoeksmethode toegepast. Volgens de 'content analysis'-methode zijn diepgaande interviews met achttien geneeskundestudenten geanalyseerd waarbij gebruik is gemaakt van een bestaand codeboek dat voortkwam uit een analyse op basis van 'grounded theory'. Tijdens de analyse werd steeds rekening gehouden met de mogelijkheid dat sommige gegevens niet in overeenstemming konden zijn met het model en werd voor deze gegevens inductieve codering toegepast. Alle drie componenten van het model, namelijk toetsfactoren, mechanismefactoren en leereffecten, bleken aanwezig te

zijn in de klinische context en alle gevonden relaties tussen deze componenten konden door het model verklaard worden. Daarnaast werd een nieuwe subcomponent aan de toetsfactoren toegevoegd, namelijk 'interactie met begeleiders'. Het model verklaarde de onderlinge relaties tussen de drie facetten van deze subcomponent (regelmatig verantwoording afleggen, consequenties op het persoonlijke vlak en de gevoelsmatige waardering voor de leeromgeving) en de eerder beschreven componenten van het model. Dit betekent dat het model geschikt is voor het analyseren en verklaren van bevindingen in andere toetsituaties dan die waarin het oorspronkelijk is ontwikkeld. De (negatieve) invloed van begeleiders op het leren van studenten kwam vooral sterk naar voren in de klinische context. Niet alle leereffecten in deze omgeving werden bepaald door ingrijpende consequenties van summatieve toetsing: persoonlijke belangen, zoals waardering en de ruimte voor zelfstandig handelen, speelden ook een rol. De resultaten geven aan dat toetsing het leren beïnvloedt als de consequenties ervan direct, concreet en aanzienlijk zijn. Het model kan gebruikt worden als instrument voor planning en diagnostiek in de onderwijspraktijk en in onderzoek van onderwijs. Er wordt ingegaan op de vraag of het model beter aansluit bij toetsing met ingrijpende consequenties dan bij summatieve toetsing.

Hoofdstuk vijf beschrijft de ontwikkeling en een eerste aanzet tot validering van een model dat kan verklaren hoe toetsing met ingrijpende consequenties het voorafgaande leren beïnvloedt. Het onderzoek is een voortzetting van het eerder gestarte valideringsproces. Om het model aan een zeer strenge test te onderwerpen werd onderzocht of de zwakste relaties tussen toetsfactoren en leereffecten in het model en de betrokken componenten getoetst konden worden en of de rol van deze subset van individuele componenten en hun onderlinge relaties bevestigd kon worden in een populatie die breder was dan, maar vergelijkbaar aan, de populatie waarvan het model oorspronkelijk was afgeleid. In een transversaal onderzoek met behulp van een speciaal daarvoor ontworpen vragenlijst zijn 21 verschillende relaties onderzocht. Elke relatie werd getoetst met behulp van situatiegebonden vragenparen ontwikkeld volgens de 'grounded-theory'-methode en gestructureerd volgens de 'key-features'- benadering. Door middel van chi-kwadraattoetsen werd per relatie bepaald of er sprake was van een significante relatie tussen toetsfactoren en leereffecten. Significantie werd bepaald met behulp van een aangepaste alfa. De frequentie van de betrokkenheid van mechanismefactoren werd berekend. Bijna alle relaties en componenten konden getoetst worden. Voor dertien van de 21 zwakste relaties werd een significante correlatie gevonden tussen toetsfactoren en leereffecten, zelfs op het significantieniveau $p < 0.00625$. De resultaten bevestigen de rol van alle acht leereffecten, alle tien mechanismefactoren en zes van de acht toetsfactoren die onderzocht zijn. Voor de subset van de zwakste relaties in het model betekent dit

een bevestiging van de rol van de meeste componenten en hun onderlinge relaties. Dit biedt gunstige vooruitzichten voor toekomstig onderzoek naar de sterkste relaties in het model. Het model blijkt dus ook van toepassing op een populatie die breder is dan, maar vergelijkbaar aan, de populatie waarop het oorspronkelijk is gebaseerd. De resultaten illustreren hoe verschillende factoren een verstorend effect kunnen hebben op interventies waarin toetsing wordt gebruikt als instrument om het leren te sturen.

Hoofdstuk zes grijpt terug op het principe van het valideren van modellen, namelijk dat interventies gebaseerd op een gevalideerd model betere resultaten moeten opleveren dan interventies gebaseerd op een zwakkere theorie. Ondanks de omvangrijke literatuur, bestaat er geen gevalideerd model van de leereffecten van toetsing en een nieuw model dient eerst gevalideerd worden voordat het als basis kan dienen voor interventies. Onderzocht is of bevindingen betreffende de generaliseerbaarheid van het model gerepliceerd konden worden in nieuwe toetsituaties. In transversaal onderzoek bij 419 studenten van twee andere universiteiten werd een speciaal voor het onderzoek ontwikkelde enquête afgenomen, bestaande uit schriftelijke situatiegebonden testen. Om het onderzoek zo stringent mogelijk te maken, werden de 21 zwakste verbanden tussen toetsfactoren en leereffecten onderzocht. Dertien hiervan bleken significant te zijn ($p < 0.00625$). De resultaten bevestigen de rol van zeven van de acht toetsfactoren, zeven van de acht leereffecten en alle tien mechanismefactoren. Een combinatie van vier mechanismefactoren (mogelijkheid tot zelfstandig handelen, doeltreffendheid van de respons, waarde van de respons en inschatting van het effect van toetsing) speelde een indirecte rol in de meeste relaties. Deze verdere bevestiging van een groot aantal van de zwakste relaties in het model biedt goede vooruitzichten voor toekomstig onderzoek naar de sterkere relaties. Hoewel de validering van het model nog niet is afgerond, komt het model door deze resultaten dichterbij het stadium waarin het een goed uitgangspunt kan zijn voor interventies.

Hoofdstuk zeven geeft een overzicht van de bevindingen en conclusies uit Hoofdstuk 2-6. Uit het onderzoek in Hoofdstuk 2 komt een volledig nieuw mechanisme naar voren dat niet alleen kan verklaren hoe studenten reageren op toetsing maar ook waarom zij zo reageren. Op basis van de bevindingen uit Hoofdstuk 3 wordt een model voorgesteld dat de leereffecten van summatieve toetsing op het voorafgaande leren kan verklaren. Op basis van de bevindingen in Hoofdstuk 4-6 wordt geconcludeerd dat het tot dusver gevonden bewijs de validiteit van het model in verschillende situaties bevestigt en dat het model beter aansluit bij toetsing met ingrijpende consequenties dan bij summatieve toetsing.

Hierna worden de vier belangrijkste onderzoeksbevindingen besproken: 1) relaties tussen toetsfactoren en leereffecten zijn complex en contextgebonden; 2) een

geheel nieuw mechanisme is gevonden dat een rol speelt in toetsing van theorie en toetsing in klinische situaties; 3) het model sluit beter aan bij toetsing met ingrijpende consequenties dan bij summatieve toetsing. Na een uitvoerige bespreking van deze vier bevindingen wordt ingegaan op het bestaande bewijs voor de validiteit van het model gezien vanuit een hiërarchie van criteria voorgesteld door Prochaska, Wright & Velicer (2008).

Vervolgens wordt ingegaan op de beperkingen van het onderzoek. De initiële onderzoeken zijn onderhevig aan de beperkingen die inherent zijn aan kwalitatief onderzoek en aan het gebruik van vrijwilligers in zowel kwalitatief als kwantitatief onderzoek. De resultaten van het kwantitatieve onderzoek zijn voornamelijk beperkt generaliseerbaar, omdat tot nu toe alleen de zwakste relaties in het model zijn onderzocht in een beperkt aantal toetsituaties.

De kracht van dit proefschrift ligt in het aangetoonde mechanisme en het ontwikkelde model, beide gebaseerd op authentieke ervaringen van studenten gedurende de volledige studieduur in een complexe, authentieke omgeving. In het kwalitatieve onderzoek is ook aandacht besteed aan de door Maxwell (2004) gesignaleerde bedreigingen van causale inferentie. Het proefschrift als geheel beschrijft niet alleen de ontwikkeling van een model ter verklaring van de leereffecten van toetsing met ingrijpende consequenties, maar tevens de eerste stappen in het valideringsproces van dit model. Hiermee levert het proefschrift een tweeledige bijdrage. In de eerste plaats biedt het model nieuwe inzichten in de relatie tussen toetsing en het voorafgaande leren door de introductie van een volledig nieuw mechanisme dat de relatie tussen toetsfactoren en leereffecten kan verklaren. In de tweede plaats, bestaat er, tenminste tot nu toe, geen ander in enige mate gevalideerd model dat de leereffecten van toetsing kan verklaren.

In de loop van dit hoofdstuk ontvouwt zich een agenda voor toekomstig onderzoek en ten slotte wordt aandacht besteed aan de implicaties van het onderzoek voor onderwijspraktijk en onderwijsontwikkeling.

References

- Krupat, E., & Dienstag, J. L. (2009). Commentary: Assessment Is an Educational Tool. *Academic Medicine*, 84(5), 548-550.
- Maxwell, J. A. (2004). Using Qualitative Methods for Causal Explanation. *Field Methods*, 16(3), 243-264.
- Prochaska, J. O., Wright, J. A., & Velicer, W. F. (2008). Evaluating Theories of Health Behavior Change: A Hierarchy of Criteria Applied to the Transtheoretical Model. *Applied Psychology*, 57(4), 561-588.

Opsomming
Summary in Afrikaans

'n Google soektog aan die begin van 2012 met die woordgreep "assessment drives learning" het 8 800 resultate opgelewer, waarvan 610 wetenskaplike artikels is. Dit is dalk 'n ietwat onwetenskaplike manier om die opsomming van 'n verhandeling te begin, maar gee tog 'n aanduiding van hoe wydverspreid hierdie opvatting is. Alhoewel die stelling wyd verkondig word en talryke navorsingsartikels oor verskeie aspekte van die onderwerp deur 'n literatuursoektog opgelewer word, is daar geen eenvormigheid oor wat presies deur "assessment drives learning" bedoel word nie. Daar is ook geen modelle waarvan die geldigheid bepaal is wat die verhouding tussen assessering en leer verklaar nie. Dit is derhalwe nie 'n verrassing dat daar in die literatuur 'n skaarste van suksesverhale is oor die doelgerigte aanwending van assessering om leer op gewenste maniere te dryf nie. Die uiteindelijke doel wat hierdie navorsing geïnspireer het is die hoop wat sedertdien deur Krupat en Dienstag (2009) verwoord is naamlik dat ons benaderings en metodes van assessering behoort te kan aanwend wat hul volle potensiaal sal gebruik ten einde ons opvoedkundige doelwitte te vervul (bl. 548). Die spesifieke kwessie waaroor hierdie verhandeling handel is die afleiding van 'n model oor die voor-assesserings leereffekte van summatielike assessering en om die geldigheid daarvan te begin bepaal.

In **Hoofstuk Een**, word die literatuur oor die verwantskap tussen assessering en leer beskou en probleme daarmee uitgelig. Die gebrek aan 'n meganisme en 'n model wat die verwantskap tussen assessering en leer verklaar, word beklemtoon. Die afbakening van die verhandeling tot die onderwerp van die voor-assesserings leereffekte van assessering word verduidelik en die viervoudige doelwitte van die verhandeling uitgespel naamlik 1) om te bepaal watter komponente van assessering watter dimensies van leer in 'n outentieke, komplekse sisteem van summatielike assessering beïnvloed het; 2) om ondersoek in te stel na watter meganisme op die spel was by die raakvlak tussen assessering en leer in hierdie assesseringstelsel; 3) om 'n teoreties begronde model te bou en 4) om stappe te neem wat die bepaling van die geldigheid van die model vorentoe sou neem.

In **Hoofstuk Twee**, word die argument aangevoer dat dit aksiomaties geword het dat assessering 'n kragtige invloed op studenteleer uitoefen, maar dat daar 'n verrassende skaarste van navorsing is oor hoe dit gebeur. 'n Studie word beskryf wat die meganisme van impak van summatielike assessering op die proses van die leer van teorie in hoër onderwys ondersoek het. Individuele, diepgaande onderhoude is met mediese studente gevoer en kwalitatief ontleed. Daar is gevind dat die impak van assessering op leer bemiddel word deur verskeie bepalers van aksie. Respondente se leergedrag is deur die volgende beïnvloed: skatting van die impak van assessering; skatting van hul leerrespons; hul persepsies van agentskap ("agency") en interpersoonlike faktore soos normatiewe oortuigings en motivering

om hieraan te voldoen. Die resultate van die studie dra by tot die karige bestaande bewyse oor die impak van assessering op leer en lei tot die voorstel van 'n meganisme om hierdie impak te verklaar.

Hoofstuk Drie beklemtoon en beskou die verbasend klein hoeveelheid navorsing wat gepubliseer is vanuit outentieke hoër onderwys- en gesondheidswetenskappe-onderrigopsette oor die aard en meganisme van die voor-assesserings leereffekte van summatiewe assessering. Hierdie studie het ondersoek ingestel na die voor-assesserings leereffekte van summatiewe assessering in teoretiese modules deur die veranderlikes op die spel in 'n veelsydige assesseringstelsel asook die verwantskappe tussen hulle na te vors. Diepgaande onderhoude is met individuele mediese studente gevoer en kwalitatief ontleed aan die hand van 'n gefundeerde teorie strategie. Respondente se leer is deur die eise van die taak en sisteemontwerp beïnvloed. Assessering het op respondente se kognitiewe prosesseringsstrategieë en metakognitiewe reguleringsaktiwiteite ingespeel. Afsonderlik bevestig ons bevindings dié van ander studies in uiteenlopende nie-mediese opsette en identifiseer dit 'n paar nuwe faktore wat in hierdie opset op die spel is. Gesamentlik, werp die bevindings van hierdie studie 'n bietjie lig op hoe 'n assesseringstelsel as geheel studenteleer oor tyd heen in 'n mediese onderrig-opset beïnvloed, die eerste studie wat so maak. Die bevindings vanuit hierdie outentieke en komplekse opset skilder 'n genuanseerde prent van hoe die interaksies tussen verskeie ingewikkelde en veelsydige faktore in 'n assesseringstelsel op mekaar inwerk om studenteleer te beïnvloed. 'n Model word voorgestel wat die bronne, meganisme en gevolge van die voor-assesserings leereffekte van summatiewe assessering koppel en wat die benutting van summatiewe assessering as 'n werktuig om leer te verhoog, sou kon versterk. Die model word op bladsy 120 van hierdie verhandeling uitgebeeld.

Hoofstuk Vier spreek die kwessie aan dat daar geen model bestaan om die leereffekte van assessering te verklaar waarvan die geldigheid bepaal is nie asook hoe hierdie leemte 'n probleem veroorsaak tydens die ontwerp van en navorsing oor assessering vir leer. Die uitdaging wat in hierdie studie aangepak word is hoe om die geldigheid te bepaal van die model wat in Hoofstuk Drie beskryf is. Die doel van die studie was dus om, as eerste stap, ondersoek in te stel of die model wel in 'n kliniese opset in werking was. Weens die kompleksiteit van die model, is 'n kwalitatiewe benadering toegepas. Data van diepgaande onderhoude met agtien mediese studente is aan inhoudsanalise onderwerp. 'n Kodeboek wat voorheen met behulp van gefundeerde teorie ontwikkel is, is gebruik. Tydens analise, is die opsie vir die moontlikheid van induktiewe kodering, as gevolg van data wat nie in die koderingsraamwerk ingepas het nie, die hele tyd voor oë gehou. Die drie komponente van die model naamlik assesseringsfaktore, meganisme faktore en

leereffekte word almal in die kliniese konteks duidelik. Verwantskappe tussen hierdie komponente kon almal aan die hand van die model verklaar word. “Interaksie met kliniese toesighouers” is as ’n nuwe sub-komponent van assesseringsfaktore identifiseer. Verwantskappe tussen die drie fasette van hierdie sub-komponent naamlik gereelde toerekenbaarheid, persoonlike gevolge en emosionele valensie van die leeromgewing en komponente van die model wat reeds voorheen beskryf is, kon aan die hand van die model verklaar word. Die model kon dus aangewend word om waarnemings in ’n assesseringskonteks andersoortig as dié waarin die model afgelei is, te ontleed en verklaar. Die (negatiewe) invloed van kliniese toesighouers op studenteleer is besonder prominent in die kliniese opset. In hierdie opset, het leereffekte nie net gespruit uit die feit dat daar met summatiewe assessering baie op die spel (“high stakes”) is nie, maar ook vanuit persoonlike gevolge bv., vir agting en agentskap (“agency”). Die resultate dui aan dat ten einde studenteleer te beïnvloed, gevolge vanuit assessering moet voortspruit wat onmiddellik, konkreet en beduidend is. Die model mag nut hê as ’n beplannings- of diagnostiese werktuig in praktyk- en navorsingsopsette. Die voorstel dat die model een van gevolgdraende assessering eerder as summatiewe assessering is, word bespreek.

In **Hoofstuk Vyf**, word die voorafgaande ontwikkeling asook poging tot die bepaling van geldigheid van die model van die voor-assesserings leereffekte van gevolgdraende assessering in oënskou geneem. Die doel van hierdie studie was om die proses van bepaling van geldigheid voor te sit en wel met ’n strawwe toets van die model. Die doelwitte was dus om te bepaal of die mins algemene verwantskappe wat bestaan tussen assesseringsfaktore en leereffekte in die model, en die komponente betrokke, toetsbaar is en of steun vir die rol van hierdie onderversameling van individuele komponente en die verwantskappe tussen hulle, gevind kon word in ’n breër maar soortgelyke populasie as dié waarin die model aanvanklik afgelei is. Een-en-twintig verskillende verwantskappe is in ’n dwarsdeursnitopname met ’n doelgemaakte vraelys ondersoek. Elke verwantskap is getoets deur gebruik te maak van pare omstandigheidstoetsitems wat aan die hand van ’n gefundeerde teorie ontwikkel en volgens die logika van die kernkenmerkebenadering gestruktureer is. Vir elke verwantskap is die chi-kwadraat statistiek gebruik om te bepaal of daar betekenisvolle assosiasies tussen assesseringsfaktore en leereffekte was. ’n Aangepaste alfa is gebruik om betekenis toe te wys. Die frekwensie van betrokkenheid van meganismefaktore is bereken. Feitlik alle verwantskappe en komponente is toetsbaar. Daar was betekenisvolle verwantskappe tussen assesseringsfaktore en leereffekte vir 13/21 van hierdie onderversameling van mins algemene verwantskappe wat ondersoek is, al is $p < 0.00625$ as betekenisvol aanvaar. Die rol van al agt leereffekte, al tien meganismefaktore en ses uit agt assesseringsfaktore wat ondersoek is, is gestaaf.

Dus, vir die onderversameling van mins algemene verwantskappe in die model, geniet die rol van meeste komponente betrokke asook die verwantskappe tussen hulle, steun. Dit hou belofte in vir toekomstige studies van die algemeenste verwantskappe in die model. Die model geld ook in 'n breër, dog soortgelyke populasie, as dit waaruit dit afgelei is. Resultate illustreer hoe 'n spektrum van faktore die resultate van intervensies wat ten doel het om assessering te gebruik om leer te beïnvloed, kan verwar.

Hoofstuk Ses herbesoek die rasionaal vir die bepaling van geldigheid van modelle naamlik dat intervensies wat gegrond is op 'n model waarvan die geldigheid bepaal is, beter uitkomst behoor op te lewer as pogings wat op swakker teorie gegrond is. Desnieteenstaande omvangryke literatuur oor die onderwerp, bestaan daar geen model, waarvan die geldigheid bepaal is, wat die leereffekte van assessering verduidelik nie. Voordat 'n model wat onlangs ontwikkel is gebruik word om intervensies toe te lig, moet die geldigheid daarvan eers bepaal word. Die doel van hierdie studie was om ondersoek daarna in te stel of voorlopige veralgemeenbaarheidsbevindings ten opsigte van die model, in nuwe assesseringskontekste herhaal kon word. Dwarsdeursnitopnames van 419 studente by twee ander universiteite is onderneem met behulp van 'n doelgemaakte vraelys bestaande uit skriftelike omstandigheidstoetsitems. Ter wille van strengheid, is die model se 21 mins algemene verwantskappe tussen assesseringsfaktore en leereffekte ondersoek. Dertien hiervan is betekenisvol ($p < 0.00625$). Die rol van sewe uit agt assesseringsfaktore, sewe uit agt leereffekte en al tien meganisme faktore wat ondersoek is, is gestaaf. Vier meganisme faktore (agentskap ["agency"], respons doelmatigheid, responswaarde en impakskatting) het saam die meeste verwantskappe bemiddel. Hierdie verdere steun vir baie van die model se mins algemene verwantskappe hou weereens belofte in vir toekomstige studies van die algemeenste verwantskappe. Alhoewel die bepaling van die geldigheid van 'n model 'n voortdurende proses is, bring hierdie resultate die model een tree nader aan 'n punt waar dit intervensies sinvol kan toelig.

Hoofstuk Sewe bied 'n oorsig van bevindings en gevolgtrekkings van Hoofstukke 2 – 6 aan. Hoofstuk 2 het 'n voorstel ten opsigte van 'n nuwe meganisme wat verklaar hoekom studente op assessering reageer op die wyses wat hulle doen en nie bloot wat hulle respons is nie, tot gevolg gehad. Bevindings vanuit Hoofstuk 3 het tot die voorstel van 'n model van die voor-assesserings leereffekte van summatiewe assessering gelei. Begrond op die bevindings wat in Hoofstukke 4 – 6 gerapporteer word, word die gevolgtrekking gemaak dat bewyse wat tot dusver ingesamel is die geldigheid van die model in verskeie opsette ondersteun en dat die model beskou kan word as een van gevolgdraende assessering eerder as van summatiewe assessering.

Vier hoof bevindings vanuit die werk word dan bespreek: 1) verwantskappe tussen assesseringsfaktore en leereffekte is kompleks en konteksgebonde; 2) 'n nuwe meganisme is op die spel in teorie- en kliniese assesseringskontekste; 3) die model geld vir gevolgdraende assessering eerder as bloot summatiewe assessering en 4) 'n model is afgelei waarvoor bewyse ten opsigte van geldigheid bestaan. Nadat daar op elk van die vier bevindings uitgebrei is, word bestaande bewyse ten opsigte van die geldigheid van die model beskou aan die hand van 'n hiërargiese stel kriteria wat deur Prochaska, Wright & Velicer (2008) voorgestel is.

Die beperkings van hierdie werk word bespreek. Vir die aanvanklike studies, sluit dit beperkings inherent aan kwalitatiewe navorsing in. Dit sluit ook beperkings ten opsigte van die gebruik van vrywilligers vir beide kwalitatiewe asook kwantitatiewe navorsing in. Vir die kwantitatiewe werk, is die potensiaal vir veralgemeenbaarheid van bevindings nog beperk, gegewe dat slegs die mins algemene verwantskappe in die model in 'n beperkte verskeidenheid assesseringskontekste tot dusver ondersoek is.

Die sterkpunte van hierdie verhandeling is dat die meganisme en die model vanaf die deurleefde ervarings van studente oor die duur van hul graadstudies heen, in 'n komplekse, outentieke opset afgelei is. Die kwalitatiewe werk het ook vele van die probleme met oorsaaklike afleidings wat deur Maxell (2004) identifiseer word, aangespreek. In die geheel gesien, stel die verhandeling 'n model voor wat die voor-assesserings leereffekte van gevolgdraende assessering verklaar en waarvoor stappe ten opsigte van die bepaling van geldigheid geneem is. Hierdie maak op twee fronte 'n bydrae. Eerstens, bied die model 'n nuwe perspektief oor hoe om assessering en voor-assesseringsleer met mekaar in verband te bring deur die koppeling van assesseringsfaktore en leereffekte met behulp van 'n nuwe meganisme. Tweedens, is die geldigheid van geen ander model wat in die verlede voorgestel is – tot op hede, in elk geval – tot enige mate bepaal nie.

'n Agenda vir toekomstige navorsing word bespreek, voordat die implikasies van die navorsing vir onderrig en onderrigontwikkeling op die einde verken word.

Verwysings

- Krupat, E., & Dienstag, J. L. (2009). Commentary: Assessment Is an Educational Tool. *Academic Medicine*, 84(5), 548-550.
- Maxwell, J. A. (2004). Using Qualitative Methods for Causal Explanation. *Field Methods*, 16(3), 243-264.
- Prochaska, J. O., Wright, J. A., & Velicer, W. F. (2008). Evaluating Theories of Health Behavior Change: A Hierarchy of Criteria Applied to the Transtheoretical Model. *Applied Psychology*, 57(4), 561-588.